

De dynamische wending in de epistemische logica

Lorenz DEMEY

Centrum voor Logica en Analytische Wijsbegeerte – KU Leuven

lorenz.demey@hiw.kuleuven.be

(Te verschijnen in *Tijdschrift voor Filosofie*.)

1. INLEIDING

De epistemische logica is het deelgebied van de filosofische logica dat zich bezighoudt met het formeel beschrijven van epistemische en doxastische noties, zoals kennis, geloof, rechtvaardiging, (subjectieve) waarschijnlijkheid, etc. Zo bestudeert men bijvoorbeeld het principe van *positieve introspectie*, dat stelt dat als een rationeel subject weet dat een of andere propositie φ waar is, zij dan ook weet dat zij weet dat φ . Dit principe wordt geformaliseerd als volgt: $K\varphi \rightarrow KK\varphi$. Hierbij is ‘K’ een unaire propositionele operator, die betekent: ‘het subject weet dat...’. Dergelijke principes zijn van groot belang vanuit filosofisch perspectief, omdat zij een grote discriminerende waarde hebben in bepaalde epistemologische debatten. Stel bijvoorbeeld dat een epistemologische theorie T1 voorspelt dat het principe van positieve introspectie geldig is, terwijl een andere epistemologische theorie T2 voorspelt dat dit principe ongeldig is (kenners van de hedendaagse epistemologie kunnen makkelijk concrete theorieën invullen voor T1 en T2). Als we nu, op basis van onafhankelijke argumenten, tot de conclusie komen dat het principe van positieve introspectie inderdaad ongeldig is, dan hebben we ook meteen een argument om de epistemologische theorie T1 te verwerpen, terwijl de theorie T2 door deze vaststelling net gecorroboereerd wordt. Op deze manier kunnen formeel-logische argumenten uit de epistemische logica een onmiddellijke impact hebben op debatten uit de epistemologie.

Een recente ontwikkeling binnen de epistemische logica is de zogenaamde *dynamische wending*. Het vroege werk in de epistemische logica focuste haast exclusief op het logische

gedrag van kennis op één specifiek ogenblik. Vandaag de dag wordt er echter meer en meer aandacht besteed aan epistemische dynamica: hoe verandert kennis doorheen de tijd, onder invloed van epistemisch relevante gebeurtenissen zoals communicatie en observatie? Veel van dit onderzoek werd en wordt verricht binnen de informatica en de speltheorie, aangezien het precies in deze disciplines is dat men veelvuldig gebruik maakt van epistemische logica bij het formeel analyseren van problemen en scenario's die expliciet epistemisch-dynamisch van aard zijn.

Een gevolg van deze ontwikkeling is dan ook dat de dynamische epistemische logica (en daarmee meteen ook een aanzienlijk gedeelte van het totale volume aan onderzoek dat vandaag de dag verricht wordt in de epistemische logica) primair gemotiveerd is door overwegingen en doeleinden uit de informatica en speltheorie. De onderzoeksresultaten die bekomen worden binnen de dynamische epistemische logica zijn van groot intern belang voor deze disciplines, maar het is twijfelachtig of zij ook relevant zijn vanuit een breder *filosofisch* perspectief. Het zal dan ook niet verbazen dat heel wat auteurs serieuze vraagtekens hebben geplaatst bij de filosofische relevantie van de dynamische wending in de epistemische logica.

Mijn hoofddoel in dit artikel bestaat erin te argumenteren dat dit scepticisme onterecht is, en dat de dynamische wending in de epistemische logica wel degelijk relevant is vanuit filosofisch perspectief. Ik zal hiertoe een onderscheid maken tussen een *sterke* en een *zwakke* interpretatie van de dynamische wending. Volgens haar zwakke interpretatie is de dynamische wending inderdaad nauwelijks filosofisch relevant, maar volgens haar sterke interpretatie past zij perfect binnen een wittgensteiniaanse opvatting over het belang van filosofie en filosofische logica. Hierdoor vindt de hedendaagse dynamische epistemische logica opnieuw aansluiting bij de expliciet filosofische motivaties die aan de grondslag lagen van het eerdere pionierswerk in de epistemische logica. In het licht van deze herwonnen filosofische relevantie heeft dit artikel als bijkomend doel een beknopte inleiding tot de dynamische epistemische logica te bieden die specifiek gericht is op een filosofisch publiek. Hierbij zal de nadruk gelegd worden op brede historische en filosofische ontwikkelingen, veeleer dan op de technische details van één of ander concreet logisch systeem.

Het artikel is opgebouwd als volgt. In Sectie 2 schets ik de historische ontwikkeling van de epistemische logica, vertrekkende bij haar begindagen in de vroege jaren '60. Hierbij zal bijzondere aandacht worden besteed aan de context waarin de dynamische wending tot stand

is gekomen, en de gevolgen hiervan voor de relatie tussen de epistemische logica en de filosofie. Ter illustratie van de dynamische wending biedt Sectie 3 een informele inleiding tot één van de belangrijkste systemen binnen deze onderzoekslijn, namelijk de logica van publieke aankondigingen. In Sectie 4 introduceer ik het onderscheid tussen de sterke en de zwakke interpretatie van de dynamische wending, en argumenteer op basis hiervan dat deze wending wel degelijk relevant kan zijn vanuit filosofisch perspectief. Om deze argumentatie te illustreren en kracht bij te zetten, presenteer ik in Sectie 5 een gevalstudie van een onderwerp waarop de dynamische epistemische logica op filosofisch vruchtbare wijze kan worden toegepast, namelijk het psychologische fenomeen ‘verrassing’. In Sectie 6, tot slot, geef ik een kernachtig overzicht van de argumentatie die doorheen dit artikel wordt ontwikkeld, en wijs ik op een aantal vragen die ons tot verdere reflectie nopen.

2. HISTORISCHE ONTWIKKELING VAN DE EPISTEMISCHE LOGICA

2.1. *De oorsprong*

Hoewel er reeds systematisch werd nagedacht over epistemische en modale operatoren in de middeleeuwen,¹ wordt het beginpunt van de hedendaagse, geformaliseerde epistemische logica meestal gesitueerd in Jaakko Hintikka’s bekende *Knowledge and Belief* uit 1962.² Hierin wordt kennis behandeld als een modale operator: we werken dus met logische formules van de vorm $K\varphi$, wat intuïtief betekent dat het (verder niet gespecificeerde) subject weet dat φ het geval is. Indien er meerdere subjecten betrokken zijn, dan worden er subject-indices toegevoegd aan de K-operatoren; zo betekent de formule $K_a\varphi \wedge \neg K_b\varphi$ bijvoorbeeld dat subject a weet dat φ het geval is, maar subject b niet.

¹ Zie bijvoorbeeld I. BOH, *Epistemic Logic in the Later Middle Ages*, London, Routledge, 1993; S. KNUUTTILA, *Modalities in Medieval Philosophy*, London, Routledge, 1993; I. BOH, ‘Four Phases of Medieval Epistemic Logic’, *Theoria* 66/2000, pp. 129-144; D. MARTENS, ‘A Late Medieval Dispute about the Conditions for Knowledge’, *Philosophical Papers* 40/2011, pp. 421-438; S. UCKELMAN, *Modalities in Medieval Logic*, doctoraatsthesis, Amsterdam, Universiteit van Amsterdam, 2011. Een meer gedetailleerd overzicht van de geschiedenis van de epistemische logica kan teruggevonden worden in P. GOCHET and P. GRIBOMONT, ‘Epistemic Logic’, in: D. GABBAY and J. WOODS (Eds.), *Handbook of the History of Logic, Volume 7*, Amsterdam, Elsevier, 2006, pp. 99-195.

² J. HINTIKKA, *Knowledge and Belief. An Introduction to the Logic of the Two Notions*, Ithaca (NY), Cornell University Press, 1962.

Deze formules worden geïnterpreteerd aan de hand van Kripk modellen. Dit zijn drietallen van de vorm $\mathbf{M} = \langle W, R, V \rangle$, waarbij W een niet-lege verzameling van mogelijke werelden is, R een binaire relatie over W , en V een valuatiefunctie die voor elke atomaire propositie specificeert in welke werelden in W deze propositie waar is. De zogenaamde ‘bereikbaarheidsrelatie’ R specificeert welke mogelijke werelden compatibel zijn met de kennis van het subject in een gegeven wereld. (Indien er verschillende subjecten zijn, dan bevat het Kripkemodel een relatie R_i voor elk subject i .) Zo betekent $(w, v) \in R$ dat wereld v compatibel is met al de kennis die het subject bezit in wereld w ; met andere woorden: gebaseerd op al de kennis die zij bezit in w , is het subject niet in staat om uit te sluiten dat v de actuele wereld zou kunnen zijn. De semantische clausule voor de K-operator ziet er als volgt uit:

$$\mathbf{M}, w \models K\varphi \quad \text{als en slechts als} \quad \text{voor alle } v \in W: \text{als } (w, v) \in R, \text{ dan } \mathbf{M}, v \models \varphi.$$

Deze clausule stelt dat het subject weet dat φ in een wereld w als en slechts als φ waar is in alle werelden die bereikbaar zijn vanuit w . Deze clausule kan uiteraard niet gezien worden als een filosofisch substantiële *definitie* van kennis, aangezien dit onmiddellijk zou resulteren in circulariteit: de kennis van het subject in een wereld w wordt ‘gedefinieerd’ in termen van de bereikbaarheidsrelatie R , die op haar beurt zelf informeel geduid wordt in termen van (compatibiliteit met) de kennis van het subject in w .

Deze concepten kunnen misschien het best verduidelijkt worden aan de hand van een kleinschalig voorbeeld. Beschouw daarom een Kripkemodel $\mathbf{M} = \langle W, R, V \rangle$ met slechts drie mogelijke werelden: $W = \{w, v, u\}$. In w is Barack Obama momenteel op bezoek in Wisconsin, in v is hij momenteel op bezoek in Vermont, en in u is hij momenteel op bezoek in Utah. We introduceren daarom drie atomaire proposities *bezoekWisconsin*, *bezoekVermont* en *bezoekUtah*, en specificeren de valuatiefunctie V als volgt:

$$V(\text{bezoekWisconsin}) = \{w\}, \quad V(\text{bezoekVermont}) = \{v\}, \quad V(\text{bezoekUtah}) = \{u\}.$$

Tot slot nemen we aan dat $(w, w) \in R$ en $(w, v) \in R$, maar $(w, u) \notin R$. Dit betekent dat het subject, op basis van al haar kennis in w , in staat is om uit te sluiten dat u de actuele wereld zou kunnen zijn, maar niet in staat is om uit te sluiten dat v de actuele wereld zou kunnen zijn.

Zij weet dus dat Obama momenteel niet Utah aan het bezoeken is, maar zij weet niet dat hij momenteel Wisconsin aan het bezoeken is (hoewel dit wel degelijk het geval is in w). Door de semantische clausule voor de K-operator toe te passen, kunnen we formeel controleren dat deze voorspellingen ook kloppen in het Kripkemodel:

$$\mathbf{M}, w \models K \neg \text{bezoekUtah},$$

$$\mathbf{M}, w \models \neg K \text{bezoekVermont},$$

$$\mathbf{M}, w \models K(\text{bezoekWisconsin} \vee \text{bezoekVermont}).$$

We hebben reeds gezien dat de semantische clausule voor de K-operator niet gebruikt kan worden als definitie van kennis. Dit betekent echter niet dat zij geheel waardeloos is vanuit filosofisch perspectief. Haar belangrijkste voordeel is dat ze ons toestaat om *correspondenties* (in een wiskundig precieze zin van het woord) te bewijzen tussen filosofische principes over kennis enerzijds en wiskundige eigenschappen van de bereikbaarheidsrelatie anderzijds. Hierdoor kunnen filosofische argumenten over de kenmerken van kennis gebaseerd worden op de wiskundige structuur van de gebruikte Kripk modellen. Typische voorbeelden van filosofische principes over kennis zijn (i) *factiviteit* (als het subject weet dat φ , dan is φ waar; formeel: $K\varphi \rightarrow \varphi$), (ii) *positieve introspectie* (als het subject weet dat φ , dan weet zij ook dat zij weet dat φ ; formeel: $K\varphi \rightarrow KK\varphi$), en (iii) *negatieve introspectie* (als het subject niet weet dat φ , dan weet zij ook dat zij niet weet dat φ ; formeel: $\neg K\varphi \rightarrow K\neg K\varphi$). Men kan aantonen dat deze filosofische principes corresponderen met respectievelijk de volgende wiskundige eigenschappen van de bereikbaarheidsrelatie R : (i) reflexiviteit ($\forall w \in W: wRw$), (ii) transitiviteit ($\forall w, v, u \in W: wRv, vRu \rightarrow wRu$), en (iii) euclidiciteit ($\forall w, v, u \in W: wRv, wRu \rightarrow vRu$).

Op basis van dergelijke correspondenties kan men argumenteren dat kennis voldoet (resp. niet voldoet) aan een filosofisch principe, door te argumenteren dat de bereikbaarheidsrelatie voldoet (resp. niet voldoet) aan de corresponderende wiskundige eigenschap. Zo kan men bijvoorbeeld tegen het principe van positieve introspectie argumenteren door aan te tonen dat de bereikbaarheidsrelatie niet transitief is.³ Beschouw hiertoe bijvoorbeeld een Kripkemodel met slechts drie mogelijke werelden, namelijk r , o en g , waarin het subject een stukje papier te zien krijgt dat resp. rood, oranje of geel gekleurd is. Het is perfect mogelijk dat het subject,

³ Zie bijvoorbeeld T. WILLIAMSON, *Knowledge and its Limits*, Oxford, Oxford UP, 2000.

omwille van haar beperkte perceptuele capaciteiten, niet in staat is om rood van oranje te onderscheiden; met andere woorden: zelfs als het stukje papier dat zij te zien krijgt de facto rood is, kan ze niet uitsluiten dat het misschien oranje zou kunnen zijn, en dus $(r,o) \in R$. Op gelijkaardige wijze is het subject niet in staat om oranje van geel te onderscheiden: zelfs als het stukje papier dat zij te zien krijgt de facto oranje is, kan ze niet uitsluiten dat het misschien geel zou kunnen zijn, en dus $(o,g) \in R$. (Elke kleur is triviaal ononderscheidbaar van zichzelf, en dus ook (r,r) , (o,o) , $(g,g) \in R$.) Uit $(r,o) \in R$ en $(o,g) \in R$ kunnen we echter *niet* automatisch besluiten dat ook $(r,g) \in R$. Het visuele contrast tussen rood en geel zou immers voldoende groot kunnen zijn opdat het subject deze twee kleuren wel degelijk van elkaar kan onderscheiden: als het stukje papier dat zij te zien krijgt de facto rood is, dan kan ze met zekerheid uitsluiten dat het geel zou kunnen zijn, en dus $(r,g) \notin R$. Hieruit blijkt dat de bereikbaarheidsrelatie R *niet* transitief is. Men kan aantonen dat dergelijke niet-transitieve bereikbaarheidsrelaties aanleiding geven tot concrete schendingen van het principe van positieve introspectie. Als het subject bijvoorbeeld een stukje papier te zien krijgt dat de facto rood is, dan weet zij dat dit stukje papier niet geel is, maar ze weet *niet* dat ze dit weet.

Op gelijkaardige wijze kan men naast een kennis- ook een geloofsoperator introduceren, en verschillende filosofische principes over de interactie tussen kennis (*knowledge*) en geloof (*belief*) bestuderen. Enkele typische voorbeelden zijn het principe van *doxastisch-epistemische introspectie* (als het subject gelooft dat φ , dan weet ze dat ze gelooft dat φ ; formeel: $B\varphi \rightarrow KB\varphi$) en het principe van *doxastisch vertrouwen* (als het subject gelooft dat φ , dan gelooft ze dat ze weet dat φ ; formeel: $B\varphi \rightarrow BK\varphi$). Men kan aantonen dat bepaalde van deze interactieprincipes (in combinatie met bepaalde principes over het gedrag van kennis en geloof afzonderlijk) ertoe leiden dat kennis en geloof samenvallen, in de zin dat de formules $K\varphi$ en $B\varphi$ logisch equivalent worden.⁴ Aangezien algemeen wordt aangenomen dat kennis en geloof verschillende propositionele attitudes zijn, kan men daarom niet tegelijkertijd al deze principes over kennis, geloof en hun interactie aanvaarden (hoe plausibel ze elk afzonderlijk misschien ook mogen lijken).

Samenvattend kan gesteld worden dat het vroege werk in de epistemische logica vaak vrij technisch van aard is, maar toch ook steeds op zeer onmiddellijke wijze verbonden is met

⁴ Zie bijvoorbeeld W. VAN DER HOEK, 'Systems for Knowledge and Belief', *Journal of Logic and Computation* 3/1993, pp. 173-195; J. HALPERN, D. SAMET and E. SEGEV, 'Defining Knowledge in Terms of Belief: The Modal Logic Perspective', *Review of Symbolic Logic* 2/2009, pp. 469-487.

filosofische vragen, concepten en debatten. Hiermee wordt niet bedoeld dat de technische resultaten op zichzelf reeds filosofische stellingnames vormen, maar wel dat zij gebruikt kunnen worden om bepaalde filosofische vragen en debatten conceptueel op te helderen.⁵

2.2. Latere ontwikkelingen

Vanaf de jaren '80 werd meer en meer onderzoek in de epistemische logica verricht door informatici en speltheoretici. Zij gebruiken concrete epistemische logica's om de vaak erg subtiele epistemische aspecten te analyseren van situaties en problemen die veelvuldig in hun vakgebied opduiken. Typische voorbeelden van dergelijke problemen uit het veld van de multi-agent systemen (artificiële intelligentie) zijn de zogenaamde 'modderige kinderen'-puzzel, de 'som en product'-puzzel en de 'Byzantijnse generaals'-puzzel; typische voorbeelden uit het veld van de protocolverificatie (cryptografie) zijn de zogenaamde 'dinerende cryptografen'-puzzel en de 'Russische kaarten'-puzzel.⁶ In de speltheorie wordt epistemische logica typisch gebruikt bij het analyseren van de epistemische voorwaarden van oplossingsconcepten voor verschillende types van spelen, zoals de achterwaartse inductie-oplossing voor spelen met perfecte informatie.⁷

⁵ Twee grote syntheseswerken uit de beginperiode van de epistemische logica zijn W. LENZEN, *Recent Work in Epistemic Logic*, Amsterdam, North-Holland, 1978 en W. LENZEN, *Glauben, Wissen und Wahrscheinlichkeit: Systeme der epistemischen Logik*, Berlin, Springer, 1980.

⁶ Zie bijvoorbeeld J. HALPERN and Y. MOSES, 'Knowledge and Common Knowledge in a Distributed Environment', *Journal of the ACM* 37/1990, pp. 549-587; H. VAN DITMARSCH, 'Het zeven-kaartenprobleem', *Nieuw Archief voor Wiskunde* 3/2002, pp. 326-332; H. VAN DITMARSCH, 'The Russian Cards Problem', *Studia Logica* 75/2003, pp. 31-62; H. VAN DITMARSCH, J. RUAN and R. VERBRUGGE, 'Sum and Product in Dynamic Epistemic Logic', *Journal of Logic and Computation* 18/2008, pp. 563-588; H. VAN DITMARSCH, J. VAN EIJCK en R. VERBRUGGE, 'Publieke werken: Freudenthal's som-en-productraadsel', *Nieuw Archief voor Wiskunde* 10/2009, pp. 126-131; H. VAN DITMARSCH, W. VAN DER HOEK and B. KOOI, *Dynamic Epistemic Logic*, Dordrecht, Springer, 2007; R. VAN DER MEYDEN, 'Two Applications of Epistemic Logic in Computer Security', in: J. VAN BENTHEM, A. GUPTA and R. PARIKH (Eds.), *Proof, Computation and Agency – Logic at the Crossroads*, Dordrecht, Springer, 2011, pp. 133-144; H. VAN DITMARSCH e.a., 'Modelling Cryptographic Keys in Dynamic Epistemic Logic with DEMO', in J. PÉREZ e.a. (Eds.), *Highlights on Practical Applications of Agents and Multi-Agent Systems*, Advances in Intelligent and Soft Computing 156, Berlin, Springer, 2012, pp. 155-162; R. PUCELLA, 'Knowledge and Security', in: H. VAN DITMARSCH, J. HALPERN, W. VAN DER HOEK and B. KOOI (Eds.), *Handbook of Logics for Knowledge and Belief*, London, College Publications, te verschijnen. Een recent en erg toegankelijk werk over deze en nog vele andere puzzels is H. VAN DITMARSCH en B. KOOI, *Honderd gevangenen en een gloeilamp*, Amsterdam, Epsilon Uitgaven, 2013.

⁷ Zie bijvoorbeeld R. AUMANN, 'Backward Induction and Common Knowledge of Rationality', *Games and Economic Behavior* 8/1995, pp. 6-19; J. HALPERN, 'Substantive Rationality and Backward Induction', *Games and Economic Behavior* 37/2001, pp. 425-435; M. KANEKO, 'Epistemic Logics and their Game Theoretic Applications: Introduction', *Economic Theory* 19/2002, pp. 7-62; B. DE BRUIN, 'Common Knowledge of Payoff Uncertainty in Games', *Synthese* 163/2008, pp. 79-97; A. BALTAG, S. SMETS and J. ZVESPER, 'Keep 'Hoping' for Rationality. A Solution to the Backward Induction Paradox', *Synthese* 169/2009, pp. 303-333.

Ter illustratie van deze toepassingen zal ik nu twee van de reeds vernoemde puzzels in wat meer detail beschrijven (zonder in te gaan op hun formeel-logische analyse). Deze puzzels kunnen misschien ver verwijderd lijken van ‘echte’ informaticatoepassingen, maar zij fungeren vooral als kleinschalige, abstracte en/of geïdealiseerde modellen om bepaalde intuïties te expliciteren. In deze zin is hun dialectische rol vergelijkbaar met die van scenario’s met namaakschuren en als zebra’s vermomde ezels in de hedendaagse epistemologie.⁸

Voorbeeld 2.1: Russische kaarten. Er zijn drie spelers: Ann, Bob en Cath. Er liggen zeven kaarten op de tafel, met de getallen 1 tot en met 7 erop. De kaarten liggen met het getal naar beneden, zodat het getal niet zichtbaar is. Ann en Bob nemen elk drie kaarten; ze letten er allebei op dat de andere spelers niet kunnen zien welke getallen er op hun kaarten staan. Cath neemt vervolgens de zevende en laatste kaart. Stel bijvoorbeeld dat Ann kaarten 1, 2 en 3 neemt, dat Bob kaarten 4, 5 en 6 neemt, en dat Cath dus de overblijvende kaart 7 krijgt. Elke speler weet nu welke kaarten hij of zij heeft, maar weet niets over de kaarten van de twee andere spelers. Kunnen Ann en Bob elkaar laten weten welke kaarten ze hebben, zonder dat Cath hierover iets te weten komt? Alle communicatie moet publiek gebeuren: alles wat Ann en Bob tegen elkaar zeggen kan gehoord worden door Cath.⁹

Voorbeeld 2.2: de modderige kinderen. Een groep kinderen heeft buiten gespeeld. Een aantal van de kinderen zijn vuil geworden: ze hebben modder op hun voorhoofd. Elk kind kan zien of de *andere* kinderen modder op hun voorhoofd hebben, maar kan niet zien of het modder op zijn *eigen* voorhoofd heeft. Vader roept de kinderen binnen en zegt: ‘Minstens één van jullie heeft modder op zijn voorhoofd’. Vervolgens vraagt hij: ‘Als je weet of je modder op je voorhoofd hebt, wil je dan nu je hand opsteken?’. Als er niets gebeurt, dan blijft vader deze laatste vraag telkens opnieuw herhalen. Wat zal er gebeuren?¹⁰

⁸ Dergelijke argumentatietechnieken werden reeds gebruikt door middeleeuwse auteurs zoals William Ockham; zie bijvoorbeeld L. DEMEY, ‘Ockham on the (In)fallibility of Intuitive Cognition’, *Logical Analysis and History of Philosophy* 17/2014, pp. 193-209.

⁹ Voor de geïnteresseerde lezer: een protocol dat tot de vermelde doelen leidt, ziet eruit als volgt. Eerst zegt Ann: ‘Mijn drie kaarten zijn 1,2,3 of 1,4,5 of 1,6,7 of 2,4,6 of 3,5,7’. Vervolgens zegt Bob: ‘Cath heeft kaart 7’. De taak van de cryptograaf bestaat er nu in om te controleren of dit protocol inderdaad ‘werkt’; met andere woorden, zij moet bewijzen dat het na deze conversatie inderdaad zo is dat Ann precies weet welke kaarten Bob heeft, dat Bob precies weet welke kaarten Ann heeft, en dat Cath nog steeds niet weet welke kaarten Ann en Bob hebben. Om een dergelijk bewijs op te stellen wordt typisch gebruik gemaakt van epistemische logica.

¹⁰ Voor de geïnteresseerde lezer: men kan bewijzen dat als er in totaal n kinderen zijn, waarvan er m modder op hun voorhoofd hebben, deze m kinderen hun hand zullen opsteken nadat vader zijn vraag m keer herhaald heeft, en dat de $n - m$ andere kinderen hun hand zullen opsteken nadat vader zijn vraag $m + 1$ keer herhaald heeft.

Samenvattend kan gesteld worden dat deze latere ontwikkelingen in de epistemische logica nog steeds vrij technisch van aard zijn, net zoals het vroege pionierswerk. De motivatie om tot deze resultaten te komen, echter, blijkt niet langer primair binnen de filosofie te liggen, maar eerder binnen de informatica en speltheorie. Op het eerste zicht lijkt het dan ook twijfelachtig dat deze resultaten, hoe interessant zij ook mogen zijn vanuit een intern computerwetenschappelijk of speltheoretisch perspectief, relevant zullen kunnen zijn vanuit een meer filosofisch perspectief.¹¹

2.3. De dynamische wending

De meeste van de hierboven genoemde scenario's uit de informatica en speltheorie hebben niet alleen betrekking op *epistemische statica* (kennis op één enkel tijdstip), maar ook op *epistemische dynamica* (het veranderen van kennis doorheen de tijd, onder invloed van epistemisch relevante gebeurtenissen zoals communicatie en observatie). In de 'modderige kinderen'-puzzel zijn er bijvoorbeeld de verschillende aankondigingen van vader, die ervoor zorgen dat de kinderen meer en meer kennis krijgen over de vraag of ze nu zelf modder op hun voorhoofd hebben of niet. Om deze scenario's adequaat te formaliseren in de epistemische logica, zijn er dan ook systemen nodig die in staat zijn om deze dynamica te representeren. De systemen uit het vroege pionierstijdperk waren echter allen volledig statisch van aard. Hintikka, bijvoorbeeld, schreef expliciet dat zijn logica *niet* geschikt was voor de analyse van situaties

waarin mensen bezig zijn nieuwe factuele informatie te verzamelen. In zo'n situaties hoeven de zinnen 'ik weet niet of p ' en [op een later tijdstip] 'ik weet dat p ' niet in tegenspraak met elkaar te zijn.¹²

De totale afwezigheid van dynamica in deze vroege epistemische logica's werd reeds in 1970 fel bekritiseerd door Dana Scott:

¹¹ Twee grote syntheseswerken wat betreft het gebruik van epistemische logica in de informatica zijn R. FAGIN, J. HALPERN, Y. MOSES and M. VARDI, *Reasoning about Knowledge*, Cambridge (MA), MIT Press, 1995 en J. MEYER and W. VAN DER HOEK, *Epistemic Logic for AI and Computer Science*, Cambridge, Cambridge UP, 1995. Twee grote syntheseswerken wat betreft het gebruik van epistemische logica in de speltheorie zijn B. DE BRUIN, *Explaining Games: The Epistemic Program in Game Theory*, Dordrecht, Springer, 2010 en A. PEREA, *Epistemic Game Theory. Reasoning and Choice*, Cambridge, Cambridge UP, 2012.

¹² "[occasions] on which people are engaged in gathering new factual information. Uttered on such an occasion, the sentences 'I don't know whether p ' and [later] 'I know that p ' need not be inconsistent.", J. HINTIKKA, *Knowledge and Belief* (pp. 7-8), mijn vertaling.

Eén van de grootste fouten van allemaal in de modale logica is volgens mij de volgende: concentreren op een systeem met slechts één modale operator. De enige manier om filosofisch significante resultaten te bereiken in [...] de epistemische logica is door deze operatoren te combineren met [bijvoorbeeld] tijdsoperatoren (want hoe zou je anders principes van verandering kunnen formuleren?).¹³

Om de puzzels waarin zij geïnteresseerd waren adequaat te kunnen formaliseren (met inbegrip van de epistemische dynamica in deze puzzels), zagen veel informatici en speltheoretici zich dus verplicht om nieuwe, *dynamische* systemen van epistemische logica te ontwikkelen. Deze dynamische epistemische logica's zijn erg succesvol gebleken bij het analyseren van vaak vrij gecompliceerde epistemisch-dynamische puzzels. Deze onderzoekslijn heeft dan ook aanleiding gegeven tot een omvangrijke literatuur, en wordt vaak aangeduid als de *dynamische wending* in de epistemische logica.¹⁴

Het is belangrijk om op te merken dat het bereik van deze dynamische wending meer omvat dan louter de epistemische logica: zij heeft ook betrekking op andere logica's van rationeel handelen, en zelfs op de logica in het algemeen. Van Benthem schrijft bijvoorbeeld:

de laatste decennia is de informatica ook invloed beginnen uitoefenen op de onderzoeksagenda van de logica. [...] er is een Dynamische Wending aan de gang in de moderne logica, waardoor activiteiten zoals inferentie, evaluatie, geloofsherziening en argumentatie centraal zijn komen te staan.¹⁵

Een goed voorbeeld van dit fenomeen vinden we in het veld van de preferentielogica, waarin

¹³ “Here is what I consider one of the biggest mistakes of all in modal logic: concentration on a system with just one modal operator. The only way to have any philosophically significant results in [...] epistemic logic is to combine those operators with [e.g.] tense operators (otherwise how can you formulate principles of change?)”, D. SCOTT, ‘Advice on Modal Logic’, in: K. LAMBERT (Ed.), *Philosophical Problems in Logic*, Dordrecht, Reidel, 1970, pp. 143-173 (p. 161), mijn vertaling.

¹⁴ Zie bijvoorbeeld A. BALTAG and L. MOSS, ‘Logics for Epistemic Programs’, *Synthese* 139/2004, pp. 165-224; H. VAN DITMARSCH, W. VAN DER HOEK and B. KOOI, *Dynamic Epistemic Logic*, Dordrecht, Springer, 2007; J. VAN BENTHEM, *Logical Dynamics of Information and Interaction*, Cambridge, Cambridge UP, 2011.

¹⁵ “over the past decades computer science has also begun to influence the research agenda of logic. [...] modern logic is undergoing a Dynamic Turn, putting activities of inference, evaluation, belief revision or argumentation at centre stage.”; J. VAN BENTHEM, ‘Logic and the Dynamics of Information’, *Minds and Machines* 13/2003, pp. 503-519 (p. 503), mijn vertaling.

de preferenties van rationeel handelende subjecten geformaliseerd worden. In één van de pionierswerken in dit veld schreef Georg Henrik von Wright bijvoorbeeld:

De preferenties die wij zullen bestuderen zijn de preferenties van een subject op één enkel moment. We sluiten dus uit [...] dat er *veranderingen* in haar preferenties kunnen optreden.¹⁶

Von Wright's beslissing om veranderingen in preferenties niet te bestuderen staat in schril contrast met de hedendaagse preferentielogica, waarin preferentiedynamica precies een centrale rol speelt.¹⁷ Hieruit blijkt duidelijk dat de dynamische wending niet alleen in de epistemische logica, maar ook in de preferentielogica heeft plaatsgevonden. Gelijkaardige opmerkingen kunnen gemaakt worden voor andere deelgebieden van de logica, zoals bijvoorbeeld de deontische logica en de kwantumlogica.¹⁸

3. DE LOGICA VAN PUBLIEKE AANKONDIGINGEN

Ter illustratie van de dynamische wending in de epistemische logica zal ik nu een eenvoudig dynamisch-epistemisch systeem bespreken, namelijk de *logica van publieke aankondigingen* (*public announcement logic*). Ondanks zijn betrekkelijke eenvoud heeft dit systeem een grote diversiteit aan toepassingen. Historisch gezien was dit bovendien het eerste dynamisch-epistemische systeem dat in detail bestudeerd werd.¹⁹ Laten we eerst een aantal typische voorbeelden van epistemische dynamica bekijken.²⁰

¹⁶ "The preferences which we shall study are a subject's [...] preferences on one occasion only. Thus we exclude [...] the possibility of *changes* in preferences."; G. H. VON WRIGHT, *The Logic of Preference*, Edinburgh, Edinburgh UP, 1963 (p. 23), mijn vertaling. Bemerkt de treffende gelijkenis tussen dit citaat en het reeds aangehaalde citaat uit Hintikka's *Knowledge and Belief* (dat bovendien slechts één jaar eerder werd gepubliceerd).

¹⁷ Zie bijvoorbeeld J. VAN BENTHEM and F. LIU, 'Dynamic Logic of Preference Upgrade', *Journal of Applied Non-Classical Logics* 17/2007, pp. 157-182; F. LIU, *Reasoning about Preference Dynamics*, Dordrecht, Springer, 2012.

¹⁸ Zie bijvoorbeeld A. ANGLBERGER, 'Dynamic Deontic Logic and Its Paradoxes', *Studia Logica* 89/2008, pp. 427-435; A. BALTAG and S. SMETS, 'Quantum Logic as a Dynamic Logic', *Synthese* 179/2011, pp. 285-306.

¹⁹ Zie J. PLAZA, 'Logics of Public Communications', in: M. EMRICH, M. PFEIFER, M. HADZIKADIC and Z. RAS (Eds.), *Proceedings of the 4th International Symposium on Methodologies for Intelligent Systems*, Oak Ridge (TN), Oak Ridge National Laboratory, 1989, pp. 201-216, herdrukt in: *Synthese* 158/2007, pp. 165-179; J. GERBRANDY and W. GROENEVELD, 'Reasoning about Information Change', *Journal of Logic, Language and Information* 6/1997, pp. 147-169.

²⁰ Deze voorbeelden zijn gebaseerd op L. DEMEY, 'Structures of Oppositions in Public Announcement Logic', in: J.-Y. BÉZIAU and D. JACQUETTE (Eds.), *Around and Beyond the Square of Opposition*, Basel, Springer, 2012, pp. 313-339.

Voorbeeld 3.1. Ann weet niet dat Parijs de hoofdstad van Frankrijk is. Tijdens een gesprek met haar beste vriend, Bob, vertelt deze laatste haar dat Parijs wel degelijk de hoofdstad van Frankrijk is. Na dit gesprek weet Ann dat Parijs de hoofdstad van Frankrijk is.

Voorbeeld 3.2. Cath weet niet dat Parijs de hoofdstad van Frankrijk is. Tijdens een gesprek met Bob vertelt deze laatste haar dat Brussel de hoofdstad van Frankrijk is (hoewel hij weet dat eigenlijk Parijs de hoofdstad van Frankrijk is). Na dit gesprek gelooft Cath dat ze weet dat Brussel de hoofdstad van Frankrijk is, maar eigenlijk weet ze dit niet.

Voorbeeld 3.3. Fred en Tom zijn concurrerende schattenjagers die net aangekomen zijn op een eiland om naar een schat te zoeken. Fred vindt een kaart waarop de exacte locatie van de schat staat aangeduid. Wat verderop heeft Tom een gesprek met een oude bewoner van het eiland, die hem informatie geeft waardoor ook Tom de exacte locatie van de schat te weten komt. De dag voordat ze de schat willen gaan opgraven, komen Fred en Tom elkaar toevallig tegen. Er hangt een vreemde spanning in de lucht tijdens hun ontmoeting...

Voorbeeld 3.1 is een duidelijk geval van een *leerproces* (een overgang van niet-weten naar weten) door middel van communicatie. In Voorbeeld 3.2 wordt Cath *misleid* door Bob: hij weet dat een zekere propositie onwaar is, en toch zegt hij tegen Cath dat ze waar is. Door deze misleiding denkt Cath onterecht dat ze nieuwe kennis heeft verworven. Voorbeeld 3.3, tot slot, illustreert het verschil tussen *private* en *publieke* communicatie: Fred en Tom hebben allebei, onafhankelijk van elkaar, kennis verworven over de locatie van de schat. Zij weten dus allebei waar de schat verborgen ligt, maar ze weten niet van elkaar dat ze dit weten. Wanneer ze elkaar toevallig ontmoeten, doet elk van hen dan ook zijn uiterste best om zijn pas opgedane kennis niet met de ander te delen (hoewel ze eigenlijk allebei reeds deze kennis bezitten).

Al deze subtiele dynamisch-epistemische effecten kunnen beschreven worden aan de hand van dynamische epistemische logica's. In het meest eenvoudige systeem, echter, namelijk de logica van publieke aankondigingen, beperken we ons tot zeer eenvoudige scenario's, zoals het volgende.

Voorbeeld 3.4. Ann en Bob zijn samen naar het televisiejournaal aan het kijken. De nieuwslezer zegt dat er morgen een treinstaking zal zijn. Ann en Bob beginnen onmiddellijk met elkaar te bespreken hoe ze morgen op hun werk kunnen geraken.

De boodschap van de nieuwslezer in dit scenario is een *publieke aankondiging*. Dit type van epistemisch-dynamische fenomenen heeft een drietal karakteristieke kenmerken. Ten eerste wordt de aankondiging gemaakt door een *externe bron* (die zelf niet expliciet geformaliseerd zal worden in de logica). In het scenario werd de aankondiging bijvoorbeeld gedaan door de ‘externe’ nieuwslezer, en niet door één van de subjecten Ann of Bob (contrasteer dit met Voorbeelden 3.1 en 3.2).²¹ Ten tweede is de aankondiging *waarheidsgetrouw*: alleen proposities die waar zijn (voorafgaand aan de aankondiging) kunnen effectief aangekondigd worden. In het scenario is het inderdaad waar dat er morgen een treinstaking zal zijn (contrasteer dit met Voorbeeld 3.2). Ten derde is de aankondiging *publiek*: ze wordt aan alle subjecten tegelijk gedaan, waardoor zij niet alleen de inhoud van de aankondiging te weten komen, maar ook dat de anderen deze inhoud nu weten. In het scenario is het niet alleen zo dat Ann en Bob elk afzonderlijk weten dat er een treinstaking zal zijn morgen, maar ook dat ze van elkaar weten dát ze dit weten. Dit blijkt duidelijk uit het feit dat ze meteen na de aankondiging van de nieuwslezer met elkaar beginnen te bespreken wat de gevolgen van deze staking zullen zijn; Ann, bijvoorbeeld, hoeft niet eerst aan Bob te zeggen dat ze zojuist vernomen heeft dat er morgen een treinstaking zal zijn: ze weet dat Bob eveneens naar de televisie aan het kijken was, en dus eveneens de aankondiging van de nieuwslezer gehoord heeft (contrasteer dit met Voorbeeld 3.3).

In de logica van publieke aankondigingen wordt deze dynamica geformaliseerd aan de hand van een *modeltransformatie*. De initiële situatie wordt geformaliseerd aan de hand van een Kripkemodel \mathbf{M} : dit model representeert de toestand van de wereld (ontische informatie) en de kennis van de subjecten hierover (epistemische informatie) vóórdat er enige aankondiging heeft plaatsgevonden. De publieke aankondiging van een formule φ transformeert dit Kripkemodel \mathbf{M} in een nieuw Kripkemodel $\mathbf{M}|\varphi$. We zullen hier niet ingaan op de precieze definitie van deze modeltransformatie, maar het is van belang om op te merken dat zij

²¹ Het onderscheid tussen ‘externe’ en ‘interne’ bronnen van aankondigingen dient overigens niet overdreven te worden. Zo zou Bobs persoonlijke aankondiging van de propositie ‘Parijs is de hoofdstad van Frankrijk’ in Voorbeeld 3.1 geformaliseerd kunnen worden als een externe aankondiging van de propositie ‘Bob weet/meent dat Parijs de hoofdstad is van Frankrijk’, aangezien beide aankondigingen exact dezelfde epistemische effecten op de subjecten lijken te hebben.

naadloos aansluit bij de statische epistemische logica: als het model \mathbf{M} een adequate representatie is van alle beschikbare ontische en epistemische informatie *voor* de publieke aankondiging van φ , dan is het nieuwe model $\mathbf{M}|\varphi$ een adequate representatie van alle beschikbare ontische en epistemische informatie *na* de publieke aankondiging van φ .

Om de effecten van deze modeltransformatie te kunnen beschrijven in de logische taal, introduceren we een nieuwe, binaire operator die van de vorm $[!\varphi]\dots$ is. Intuïtief betekent de formule $[!\varphi]\psi$ dat ψ het geval zal zijn na een publieke aankondiging van φ . De formeel-semantische clausule voor deze dynamische operator ziet er als volgt uit:

$$\mathbf{M}, w \models [!\varphi]\psi \quad \text{als en slechts als} \quad \text{als } \mathbf{M}, w \models \varphi, \text{ dan } \mathbf{M}|\varphi, w \models \psi.$$

Deze clausule definieert de waarheid van $[!\varphi]\psi$ in (een wereld w in) het Kripkemodel \mathbf{M} in termen van de waarheid van de deelformule ψ in (diezelfde wereld w in) het nieuwe Kripkemodel $\mathbf{M}|\varphi$. De conditionele structuur van deze clausule reflecteert de waarheidsgetrouwheid van publieke aankondigingen: het ‘nieuwe’ model $\mathbf{M}|\varphi$ speelt enkel een rol als de aangekondigde formule φ waar is in (de wereld w in) het ‘huidige’ model \mathbf{M} . Aan de hand van deze operatoren kunnen scenario’s zoals Voorbeeld 3.4 op een erg compacte en natuurlijke wijze beschreven worden in de formele taal. Als het Kripkemodel \mathbf{M} een adequate representatie is van de situatie voorafgaand aan de aankondiging door de nieuwslezer, w de actuele wereld is, en p de propositie dat er morgen een treinstaking zal plaatsvinden, dan geldt:

$$\mathbf{M}, w \models \neg K_a p \wedge \neg K_b p \wedge [!p](K_a p \wedge K_b p \wedge K_a K_b p \wedge K_b K_a p).$$

Deze formule stelt dat initieel (voor de aankondiging door de nieuwslezer) Ann en Bob niet weten dat er morgen een treinstaking zal plaatsvinden, maar dat zij dit na de aankondiging door de nieuwslezer wél weten, en bovendien ook weten van elkaar dát ze het weten.

4. STERKE EN ZWAKKE INTERPRETATIES VAN DE DYNAMISCHE WENDING

In deze sectie maak ik een onderscheid tussen een sterke en een zwakke interpretatie van de dynamische wending in de epistemische logica. Volgens de zwakke interpretatie lijkt de dynamische wending nauwelijks enige filosofische relevantie te bezitten, maar volgens de sterke interpretatie zal zij wel degelijk van grote filosofische waarde blijken te zijn.

De *zwakke interpretatie* van de dynamische wending leunt zeer dicht aan bij de historische feiten over de ontwikkeling van de epistemische logica, zoals deze gepresenteerd werden in Sectie 2 van dit artikel. Volgens deze interpretatie bestaat het *primaire*, of zelfs het *enige*, doel van dynamische epistemische logica's erin om scenario's zoals de 'Russische kaarten'-puzzel en de 'modderige kinderen'-puzzel formeel te analyseren. Deze scenario's zijn expliciet dynamisch van aard: het is bijvoorbeeld duidelijk dat de communicatie van de spelers in de 'Russische kaarten'-puzzel (zie Voorbeeld 2.1 en voetnoot 9) en de verschillende aankondigingen van vader en de kinderen in de 'modderige kinderen'-puzzel (zie Voorbeeld 2.2 en voetnoot 10) adequaat geformaliseerd kunnen worden in de logica van publieke aankondigingen (zie Sectie 3). We hebben ook reeds gezien dat het precies dit soort scenario's was dat de concrete motivatie vormde voor de ontwikkeling van dynamisch-epistemische systemen, en aldus geleid heeft tot de dynamische wending in de epistemische logica.

Volgens de zwakke interpretatie van de dynamische wending blijft de relevantie van de dynamische epistemische logica dus beperkt tot het analyseren van scenario's en puzzels die de oorspronkelijke motivatie vormden voor haar ontwikkeling. We hebben reeds in Sectie 2 van dit artikel gezien dat de overgrote meerderheid van deze puzzels afkomstig is uit de informatica en de speltheorie.²² Hoewel deze analyses erg succesvol zijn gebleken vanuit intern computerwetenschappelijk of speltheoretisch perspectief, zal het wellicht niet verbazen dat zij niet meteen relevant zijn vanuit *filosofisch* perspectief. Ter illustratie: dynamische epistemische logica's zijn erg nuttig gebleken bij het analyseren van de epistemische voorwaarden van cryptografische protocollen (men denke aan Voorbeeld 2.1 en voetnoot 9) en vormen volgens sommige auteurs dan ook een waardevol instrument in de hedendaagse cryptografie, maar het is niet meteen duidelijk hoe de vele resultaten die hiermee bereikt zijn

²² Er kan niet ontkend worden dat er wel degelijk toepassingen van dynamische epistemische logica's bestaan die betrekking hebben op expliciet dynamische problemen uit de *filosofie* (en dus niet uit de informatica of speltheorie). Een goed voorbeeld is de formalisatie van de middeleeuwse theorie van *obligationes*; zie bijvoorbeeld S. UCKELMAN, 'Deceit and Infeasible Knowledge: the Case of Dubitatio', *Journal of Applied Non-Classical Logics* 21/2011, pp. 503-519; S. UCKELMAN, 'A Dynamic Epistemic Logic Approach to Modeling Obligationes', in: D. GROSSI, S. MINICA, B. RODENHÄUSER and S. SMETS (Eds.), *Logic and Interactive Rationality Yearbook 2010*, Amsterdam, Institute for Logic, Language and Computation, 2011, pp. 147-172; S. UCKELMAN, 'Medieval *Disputationes de obligationibus* as Formal Dialogue Systems', *Argumentation* 27/2013, pp. 143-166. Deze filosofische toepassingen zijn echter veel schaarser dan toepassingen binnen de informatica en de speltheorie. Bovendien zijn deze filosofische toepassingen van veel recentere datum, en behoren zij dus zeker niet tot de toepassingen die de oorspronkelijke motivatie vormden voor de ontwikkeling van de dynamische epistemische logica.

van nut zouden kunnen zijn bij het conceptueel ophelderen van filosofische vragen en debatten.

De zwakke interpretatie van de dynamische wending leidt dus tot een zekere mate van scepticisme over de filosofische relevantie van deze wending. Sporen van dit scepticisme – dat trouwens ook van toepassing is op andere logica's die historisch gezien primair gemotiveerd waren door toepassingen in de informatica en/of de speltheorie – kunnen teruggevonden worden in de houding van verschillende filosofen tegenover deze logica's. In een recent overzichtswerk over de geschiedenis van de filosofische logica lezen we bijvoorbeeld:

de epistemische logica werd opnieuw erg populair omwille van de interesse van informatici in de jaren 1980. Een *ontwikkeling die vanuit filosofisch perspectief interessanter is* dan formele systemen is gelegen in de studie naar objectieve en subjectgebonden kwantificatie.²³

Laten we nu overgaan op de *sterke interpretatie* van de dynamische wending. Deze interpretatie benadrukt dat dynamische epistemische logica's niet alleen gebruikt kunnen worden om expliciet dynamische scenario's en puzzels te modelleren, maar dat zij ook een nieuw licht kunnen werpen op scenario's, noties en stellingen die *prima facie* helemaal statisch van aard zijn. Het is immers best denkbaar dat dit statische voorkomen slechts een façade is, waarachter een heleboel dynamische fenomenen verscholen kunnen liggen.²⁴ Dynamische epistemische logica's zullen ons toestaan om deze verborgen dynamica bloot te leggen en volledig expliciet te maken, en daardoor meer adequate conceptuele analyses van deze scenario's, noties en stellingen te bereiken. Bij van Benthem lezen we bijvoorbeeld:

de motivatie voor standaardlogica's bevat vaak procedurele elementen, die ook

²³ “epistemic logic again became very popular because of the interest of computer scientists in the 1980s. A philosophically more interesting development than formal calculi may be in the speculation concerning objective and subject-bound quantification.”; T. KORTE, A. MAUNU and T. AHO, ‘Modal Logic from Kant to Possible World Semantics’, in: L. HAAPARANTA (Ed.), *The Development of Modern Logic*, Oxford, Oxford UP, 2009, pp. 516-550 (p. 27), mijn vertaling, mijn cursivering.

²⁴ Dit kan ook in verband gebracht worden met het semantische onderscheid tussen *processen* en *producten*. Veel woorden kunnen een afgewerkt product aanduiden (statische notie), maar eveneens het proces waarmee dat product tot stand is gekomen (dynamische notie). Een woord als ‘ontdekking’ kan bijvoorbeeld slaan op het eigenlijke ‘proces van ontdekken’, maar ook op het product dat voortvloeit uit dat proces, nl. datgene wat ontdekt is (met dank aan een anonieme referent voor dit mooie voorbeeld). Zie verder J. VAN BENTHEM, *Logical Dynamics of Information and Interaction*, Cambridge, Cambridge UP, 2011 (p. 1).

vaak aan bod komen in logicahandboeken – en deze impliciete dynamica kan expliciet gemaakt worden.²⁵

Dit soort van conceptuele opheldering is één van de belangrijkste manieren waarop logica van nut kan zijn voor de filosofie. Bij het bespreken van de filosofische relevantie van de vroege resultaten uit de epistemische logica in Sectie 2.1 van dit artikel, heb ik er bijvoorbeeld reeds op gewezen dat deze resultaten gebruikt kunnen worden om verschillende filosofische vragen en concepten op te helderen. Dit perspectief sluit bovendien nauw aan bij Wittgensteins *Tractatus*, waarin gesteld wordt dat niet alleen de logica, maar zelfs de filosofie als geheel gericht zou moeten zijn op conceptuele opheldering:

Filosofie is gericht op de logische opheldering van gedachten. Filosofie is geen verzameling leerstellingen, maar een activiteit. Een filosofisch werk bestaat essentieel uit ophelderingen.²⁶

Deze gedachte komt ook terug in *Must Do Better*, Timothy Williamsons zelfverklaarde sermoeen over de huidige staat van de filosofie, wanneer hij het heeft over het belang van de logica (en de wiskunde in het algemeen) voor de filosofie:

De filosofie kan nooit gereduceerd worden tot de wiskunde. Maar we kunnen vaak wiskundige modellen opstellen voor fragmenten van de filosofie, en telkens als dit mogelijk is, moeten we het ook doen. Ongetwijfeld zullen deze modellen vaak wilde idealisaties bevatten, maar de poging om een niet-triviaal model op te stellen van een filosofisch idee brengt soms reeds een verborgen structurele incoherentie in dat idee zelf aan het licht.²⁷

Volgens haar sterke interpretatie is de dynamische wending in de epistemische logica dus wel degelijk filosofisch relevant, aangezien zij kan bijdragen tot de conceptuele opheldering van

²⁵ “the motivation for standard logics often contains procedural elements present in textbook presentations – and one can make this implicit dynamics explicit”; J. VAN BENTHEM, *Exploring Logical Dynamics*, Stanford (CA), CSLI Publications, 1996 (p. 17), mijn vertaling.

²⁶ “Philosophy aims at the logical clarification of thoughts. Philosophy is not a body of doctrine but an activity. A philosophical work consists essentially of elucidations.”; L. WITTGENSTEIN, *Tractatus Logico-Philosophicus*, London, Kegan Paul, 1922 (propositie 4.112), mijn vertaling.

²⁷ “Philosophy can never be reduced to mathematics. But we can often produce mathematical models of fragments of philosophy and, when we can, we should. No doubt the models usually involve wild idealizations [but] the attempt to construct a non-trivial model [of a philosophical idea sometimes] reveals a hidden structural incoherence in the idea itself”; T. WILLIAMSON, *The Philosophy of Philosophy*, London, Blackwell, 2007 (p. 291), mijn vertaling.

bepaalde noties (voorbeelden volgen later), wat volgens auteurs zoals Wittgenstein en Williamson tot de kerntaken van de filosofie behoort. De hogere mate van *conceptuele hygiëne* die hierdoor bereikt kan worden (adequater conceptuele analyses van de noties in kwestie) is reeds filosofisch waardevol op zichzelf, maar zij leidt bovendien vaak tot verrassende bijkomende voordelen, zoals een hogere empirische dekkingsgraad, of een radicale herijking van het relatieve belang van bepaalde noties.

Het succes van deze sterke interpretatie is echter in hoge mate afhankelijk van de aanname dat er inderdaad scenario's, noties en stellingen bestaan die *prima facie* helemaal statisch van aard zijn, maar toch verborgen dynamische aspecten blijken te hebben, en daarom beter in een dynamische epistemische logica geanalyseerd kunnen worden. In de literatuur van de laatste jaren kunnen reeds verschillende treffende voorbeelden hiervan teruggevonden worden, bijvoorbeeld met betrekking tot (i) de lockeaanse stelling in de epistemologie,²⁸ (ii) Aumanns befaamde *agreeing to disagree*-stelling en het relatieve belang van gemeenschappelijke kennis (*common knowledge*) voor deze stelling,²⁹ (iii) het probleem van logische alwetendheid,³⁰ en (iv) de Fitch-paradox over ongekende en onkenbare waarheden.³¹ Om mijn argumentatie te illustreren en verder kracht bij te zetten, zal ik in de volgende sectie nog een extra voorbeeld aan dit lijstje toevoegen, en ook expliciet aangeven hoe de verschillende voordelen van een dynamische analyse in dit concrete voorbeeld tot uiting komen.

5. GEVALSSTUDIE: EEN DYNAMISCH PERSPECTIEF OP VERRASSING

Verrassing³² is een wijdverspreid fenomeen in het alledaagse leven: mensen zijn vaak verrast, bijvoorbeeld door een onverwachte lichtflits, of – minder sensationeel – omdat alle melk in de

²⁸ Zie bijvoorbeeld L. DEMEY, 'Contemporary Epistemic Logic and the Lockean Thesis', *Foundations of Science* 18/2013, pp. 599-610.

²⁹ Zie bijvoorbeeld C. DÉGREMONT and O. ROY, 'Agreement Theorems in Dynamic Epistemic Logic', *Journal of Philosophical Logic* 41/2012, pp. 735-764; L. DEMEY, 'Agreeing to Disagree in Probabilistic Dynamic Epistemic Logic', *Synthese* 191/2014, pp. 409-438.

³⁰ Zie bijvoorbeeld H. DUC, 'Reasoning about Rational, but not Logically Omniscient, Agents', *Journal of Logic and Computation* 7/1997, pp. 633-648; T. ÅGOTNES en N. ALECHINA, 'The Dynamics of Syntactic Knowledge', *Journal of Logic and Computation* 17/2007, pp. 83-116; F. VELÁZQUEZ-QUESADA, 'Inference and Update', *Synthese* 169/2009, pp. 283-300; J. VAN BENTHEM en F. VELÁZQUEZ-QUESADA, 'The Dynamics of Awareness', *Synthese* 177/2010, pp. 5-27.

³¹ Zie bijvoorbeeld J. VAN BENTHEM, 'What One May Come to Know', *Analysis* 64/2004, pp. 95-105; P. BALBIANI e.a., '"Knowable" as "Known After an Announcement"', *Review of Symbolic Logic* 1/2008, pp. 305-334; J. VAN BENTHEM, 'Actions that Make us Know', in: J. SALERNO (Ed.), *New Essays on the Knowability Paradox*, Oxford, Oxford UP, 2009, pp. 129-146.

³² Deze sectie is deels gebaseerd op L. DEMEY, 'The Dynamics of Surprise', *Logique et Analyse*, te verschijnen. In dat artikel wordt er (veel) dieper ingegaan op de empirisch-psychologische theorievorming over verrassing, de

buurtwinkel uitverkocht blijkt te zijn (de buurtwinkel is immers meestal goed bevoorraad). Verrassing wordt dan ook intensief bestudeerd door psychologen, vanuit cognitieve, sociale, ontwikkelingspsychologische en pedagogische perspectieven. Maar ook binnen de logica en de filosofie speelt verrassing een belangrijke rol. Zo is er bijvoorbeeld de befaamde ‘paradox van de verrassingstest’ (*surprise examination paradox*),³³ en Wittgensteins anti-psychologistische dictum dat er in de logica geen verrassingen mogelijk zijn.³⁴ Verder deed bijvoorbeeld Peirce een beroep op verrassing bij het bespreken van zijn notie van abductief redeneren, terwijl Davidson de relatie tussen verrassing en verwachting bestudeerd heeft.³⁵

Binnen de psychologie bestaat er een vrij grote consensus dat verrassing een emotie is die met verschillende graden van intensiteit kan voorkomen, en dat deze intensiteit monotoon afhankelijk is van de mate van onverwachtheid van de verrassende gebeurtenis; met andere woorden: hoe onverwachter de verrassende gebeurtenis, hoe sterker de verrassing die zij teweegbrengt. Op basis van dit inzicht kan men een verklaring bieden voor verschillende experimenteel vastgestelde eigenschappen en patronen, zoals bijvoorbeeld de relatie tussen verrassing en *belief revision* (mensen die verrast zijn over iets gaan vaak nadien hun mening over dit onderwerp herzien) en het *vluchtige* karakter van verrassing (als een verrassende gebeurtenis steeds herhaald wordt, dan verliest zij haar verrassende effect).

In de recente literatuur zijn er een aantal pogingen geweest om deze psychologische theorieën over verrassing in een formeel-logisch systeem te gieten.³⁶ Deze systemen zijn echter te

klassieke (niet-dynamische) formalisering van verrassing, de ontwikkeling van het nieuwe (dynamische) systeem, zijn logische en metalogische eigenschappen, de vergelijking tussen deze verschillende systemen, etc.

³³ Zie bijvoorbeeld T. CHOW, ‘The Surprise Examination or Unexpected Hanging Paradox’, *American Mathematical Monthly* 105/1998, pp. 41-51; R. SORENSEN, ‘Epistemic Paradoxes’, in: E. ZALTA (Ed.), *Stanford Encyclopedia of Philosophy*, Stanford (CA), CSLI, 2011. De analyse van deze paradox staat ook reeds geruime tijd op de onderzoeksagenda binnen de dynamische epistemische logica; zie bijvoorbeeld H. VAN DITMARSCH en B. KOOI, ‘Een analyse van de Hangman paradox in dynamische epistemische logica’, *Algemeen Nederlands Tijdschrift voor Wijsbegeerte* 97/2005, pp. 16-30; H. VAN DITMARSCH en B. KOOI, ‘The Secret of My Success’, *Synthese* 151/2006, pp. 201-232; J. GERBRANDY, ‘The Surprise Examination in Dynamic Epistemic Logic’, *Synthese* 155/2007, pp. 21-33.

³⁴ “there can never be surprises in logic”; L. WITTGENSTEIN, *Tractatus Logico-Philosophicus*, London, Kegan Paul, 1922, propositie 6.1251.

³⁵ C. S. PEIRCE, *The Collected Papers of Charles Sanders Peirce, Vol. V: Pragmatism and Pragmaticism*, Cambridge (MA), Harvard University Press, 1934, §189; D. DAVIDSON, ‘Rational Animals’, *Dialectica* 36/1982, pp. 317-327.

³⁶ Zie bijvoorbeeld L. MACEDO and A. CARDOSO, ‘Modelling Forms of Surprise in an Artificial Agent’, in: J. MOORE and K. STENNING (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society*, Edinburgh, Lawrence Erlbaum, pp. 588-593; L. MACEDO, A. CARDOSO, R. REISENZEIN, E. LORINI and C. CASTELFRANCHI, ‘Artificial Surprise’, in: J. VALLVERDÚ and D. CASACUBERTA (Eds.), *Handbook of Research on Synthetic Emotions and Sociable Robotics: New Applications in Affective Computing and AI*, Hershey (PA), IGI Global, 2009, pp. 267-291; E. LORINI and C. CASTELFRANCHI, ‘The Unexpected Aspects of Surprise’,

statisch van aard: zij focussen op verrassing als ‘stabiele emotie’, zonder in kaart te brengen hoe deze emotie systematisch afhangt van de gebeurtenis die haar veroorzaakt heeft (nl. de verrassende gebeurtenis). Laten we $P(\varphi)$ gebruiken om te spreken over de waarschijnlijkheid die het subject toekent aan de propositie φ , en $S(\varphi)$ om te spreken over de graad van intensiteit waarmee het subject verrast is over φ . Een bijzonder eenvoudig axioma uit één van de voornoemde logische systemen ziet er dan als volgt uit:

$$S(\varphi) = 1 - P(\varphi).$$

Dit kernaxioma lijkt een rechtstreekse formalisering te zijn van de gedachte dat verrassingsintensiteit monotoon gecorreleerd is met mate van onverwachtheid (oftewel: omgekeerd gecorreleerd met mate van verwachting). Toch is er op conceptueel vlak iets grondig fout met dit axioma. De linkerkant van de identiteit heeft betrekking op (de intensiteit van) verrassing over φ , meer bepaald, de verrassing die door het subject ervaren wordt *nadat* de verrassende gebeurtenis heeft plaats gevonden (immers, voorafgaand aan deze gebeurtenis ervaart het subject helemaal geen verrassing). De rechterkant van de identiteit heeft echter betrekking op de subjectieve waarschijnlijkheid die het subject aan φ toekent *vóórdat* de verrassende gebeurtenis heeft plaats gevonden (immers, na deze gebeurtenis zal het subject haar waarschijnlijkheid typisch herzien hebben in het licht van de gebeurtenis). Het bovenstaande axioma legt dus een verband tussen een cognitieve/emotionele attitude *na* de verrassende gebeurtenis en een cognitieve/emotionele attitude *voor* de verrassende gebeurtenis,³⁷ maar de verrassende gebeurtenis zélf – die als ‘brug’ zou moeten fungeren tussen beide attitudes – wordt nergens expliciet gemaakt in de formalisering.

Door over te schakelen op een dynamische logica wordt het mogelijk om niet alleen de twee cognitieve/emotionele attitudes van het subject uit te drukken, maar ook de verrassende gebeurtenis zélf. De logica van publieke aankondigingen, die geïntroduceerd werd in Sectie 3 van dit artikel, volstaat reeds voor dit doeleinde. Laten we de verrassende gebeurtenis

International Journal of Pattern Recognition and Artificial Intelligence 20/2006, pp. 817-833; E. LORINI and C. CASTELFRANCHI, ‘The Cognitive Structure of Surprise: Looking for Basic Principles’, *Topoi* 26/2007, pp. 133-149.

³⁷ Dit is de geïntendeerde betekenis van het bovenstaande axioma. Een andere betekenis ontstaat wanneer beide attitudes (aan linker- en rechterzijde van de identiteit) geïnterpreteerd worden als *voorafgaand aan* de verrassende gebeurtenis; nog een andere betekenis ontstaat wanneer beide attitudes geïnterpreteerd worden als *volgend op* de verrassende gebeurtenis. Het is echter makkelijk in te zien dat deze twee alternatieve interpretaties nonsensicaal zijn, en absoluut niet overeenkomen met de informele gedachte dat verrassingsintensiteit monotoon gecorreleerd is met mate van onverwachtheid.

modelleren als een publieke aankondiging.³⁸ Het ‘tijdstip’ waarop het subject een bepaalde cognitieve/emotionele propositionele attitude heeft, kan nu simpelweg afgelezen worden aan de syntactische vorm van de formule: als de betreffende operator *binnen* het bereik van de publieke aankondigingsoperator staat, dan gaat het om een attitude *nadat* de verrassende gebeurtenis heeft plaatsgevonden; als de operator *buiten* dit bereik staat, dan gaat het om een attitude *voordat* de verrassende gebeurtenis heeft plaatsgevonden. Bijvoorbeeld: $P(\varphi) = c$ betekent dat het subject waarschijnlijkheid c toekent aan φ *voorafgaand aan* de verrassende gebeurtenis, terwijl $[!\varphi]P(\varphi) = c$ betekent dat het subject waarschijnlijkheid c toekent aan φ *volgend op* de verrassende gebeurtenis.

We zijn nu in staat om het bovenstaande kernaxioma op een dynamische, en dus meer adequate manier te herformuleren:

$$[!\varphi]S(\varphi) = c \leftrightarrow P(\varphi) = 1 - c.$$

Informeel zegt dit axioma dat het subject verbaasd zal zijn met intensiteit c over φ *na* de verrassende gebeurtenis als en slechts als zij waarschijnlijkheid $1 - c$ toekende aan φ *voor* de verrassende gebeurtenis.³⁹ Zoals we hierboven reeds zagen, is dit ook de *geïntendeerde* betekenis van de vorige (statische) formulering van dit axioma, maar deze formulering heeft ook andere, *niet-geïntendeerde* betekenissen (zie voetnoot 37). Omdat de nieuwe (dynamische) herformulering ook de verrassende gebeurtenis expliciet weergeeft (door middel van de $[!\varphi]$ -operator), verzekert zij dat enkel de geïntendeerde betekenis overblijft, en dat de andere, niet-geïntendeerde betekenissen niet langer beschikbaar zijn. Op deze manier bereikt de dynamische herformulering van het axioma een veel grotere mate van conceptuele hygiëne dan de oorspronkelijke, statische formulering.

³⁸ Een verrassende gebeurtenis zoals een onverwachte lichtflits wordt dus gezien als een publieke aankondiging van de propositie ‘er doet zich een lichtflits voor op tijdstip T en locatie L ’. Dit lijkt misschien enigszins bizar, maar het is van belang om de abstract-logische notie van ‘publieke aankondiging’ niet te verengen tot een louter linguïstisch-communicatief proces (‘een aankondiging die door iemand uitgesproken wordt’); zij kan evengoed geïnterpreteerd worden in termen van zintuiglijke observatie, leerprocessen, etc. Zie ook J. VAN BENTHEM, J. GERBRANDY and B. KOOI, ‘Dynamic Update with Probabilities’, *Studia Logica* 93/2009, pp. 67-96 (p. 71).

³⁹ Een meer algemene en nauwkeurige versie van de dynamische herformulering van het kernaxioma ziet er als volgt uit: $[!\varphi]S(\psi) = c \leftrightarrow [\varphi \rightarrow P(\langle !\varphi \rangle \psi) = 1 - c]$. De verschillen tussen deze versie en degene die in de hoofdttekst wordt gegeven, zijn echter vooral van technische aard (zie ook voetnoot 40), en zijn verder van weinig belang voor het huidige betoog.

Het is mogelijk om deze dynamische herformulering van het kernaxioma over verrassing verder uit te werken tot een volwaardig dynamisch logisch systeem met een modeltheoretische semantiek en een correcte en volledige axiomatisering. We zullen in dit artikel niet ingaan op de technische details van deze uitwerking, maar wél nog even stilstaan bij een aantal stellingen die binnen dit systeem bewezen kunnen worden. Deze stellingen vormen immers goede illustraties van de voordelen van een dynamische analyse van verrassing, en onderstrepen aldus nogmaals het belang van de sterke interpretatie van de dynamische wending in de epistemische logica.

Laten we eerst kijken naar de expressiekracht van het systeem. Deze komt het best tot uiting bij de analyse van complexe scenario's en puzzels. Beschouw bijvoorbeeld de volgende formule:

$$\begin{aligned}
 p \quad & \wedge \quad \neg Kp \quad \wedge \quad P(p) = 0.05 \quad \wedge \quad S(p) = 0 \\
 & \wedge \quad [!p] (\quad Kp \quad \wedge \quad P(p) = 1 \quad \wedge \quad S(p) = 0.95 \quad) \\
 & \wedge \quad [!p][!p] (Kp \quad \wedge \quad P(p) = 1 \quad \wedge \quad S(p) = 0 \quad).
 \end{aligned}$$

Als p staat voor ‘het is aan het sneeuwen’, dan beschrijft deze formule het volgende scenario:

Hoewel het momenteel aan het sneeuwen is, weet Jan niet dat dit het geval is. Jan is er bovendien sterk van overtuigd dat het momenteel niet aan het sneeuwen is [bijvoorbeeld omdat het april is, en Jan weet dat sneeuw in april vrij zeldzaam is], en acht het dus slechts 5% waarschijnlijk dat het momenteel aan het sneeuwen is. Aangezien Jan nog niet te weten is gekomen dat het aan het sneeuwen is, ervaart hij hier momenteel ook geen enkele verrassing over. Vervolgens kijkt hij uit het raam, en ziet dat het aan het sneeuwen is. Dit heeft een aantal cognitieve en emotionele gevolgen: op cognitief vlak zal Jan nu wél weten dat het aan het sneeuwen is, en zijn waarschijnlijkheid van sneeuw herzien van 5% naar de volle 100%; op emotioneel vlak zal hij bovendien erg verrast zijn dat het aan het sneeuwen is (omdat hij dit aanvankelijk slechts 5% waarschijnlijk achtte). Vijf minuutjes later kijkt Jan opnieuw door het raam, en ziet dat het nog steeds aan het sneeuwen is. Dit heeft geen cognitieve gevolgen meer (Jans kennis en waarschijnlijkheden blijven ongewijzigd sinds de eerste

keer dat hij uit het raam keek), en op emotioneel vlak is hij nu niet langer verrast (aangezien het vijf minuten geleden aan het sneeuwen was, is het niet echt verbazingwekkend dat het nu nog steeds aan het sneeuwen is).

In dit scenario kunnen drie relevante ‘fases’ onderscheiden worden: (i) voordat Jan naar buiten gekeken heeft, (ii) nadat Jan een eerste keer naar buiten gekeken heeft, en (iii) nadat Jan een tweede keer naar buiten gekeken heeft. Telkens als Jan naar buiten kijkt, is er dus een overgang van de ene fase naar de volgende. Deze temporeel gelaagde structuur is ook duidelijk aanwezig in de bovenstaande formule: elke fase wordt gekarakteriseerd door een Kp -, een $P(p)$ - en een $S(p)$ -conjunct, en de overgangen tussen de fases worden gemodelleerd aan de hand van dynamische $[!p]$ -operatoren. Het zijn dus precies deze dynamische operatoren die ons in staat stellen om de opeenvolging van fases formeel uit te drukken, waardoor de connectie tussen het intuïtieve scenario en de logische modellering ervan volledig transparant wordt.

Verder kan ook de volgende stelling bewezen worden in het dynamische systeem:

$$\models (P(\varphi) = a \wedge [!p]P(\varphi) = b \wedge a \neq b) \rightarrow [!p]S(\varphi) > 0.$$

Deze stelling zegt dat indien er zich een gebeurtenis voordoet die als effect heeft dat het subject haar waarschijnlijkheden herzielt (de aan φ toegekende waarschijnlijkheid verandert van a in b), dit stevast gepaard zal gaan met enige mate van verrassing (met strikt positieve intensiteit). Met andere woorden: verrassing is een noodzakelijke voorwaarde voor *belief revision*. Deze stelling vormt dus een belangrijke connectie met de psychologische theorievorming, waarin eveneens gesteld wordt dat verrassing onlosmakelijk verbonden is met *belief revision*.

Tot slot bekijken we nog één stelling die bewezen kan worden in het dynamische systeem: als φ succesvol⁴⁰ is, dan geldt er dat

⁴⁰ We zagen in Sectie 3 reeds dat een formule φ slechts aangekondigd kan worden als zij waar is *voor* de aankondiging. Het lijkt natuurlijk om aan te nemen dat φ nog steeds waar zal zijn *na* de aankondiging. Omdat φ echter epistemische operatoren kan bevatten, klopt dit niet in het algemeen: het kan zijn dat φ , simpelweg door aangekondigd te worden, onwaar wordt. Een bekend voorbeeld hiervan is de Moore-zin $p \wedge \neg Kp$. Formules waarvoor er zo geen ‘zelf-falsifiërende’ effecten optreden, worden *succesvol* genoemd; met andere woorden: φ is succesvol als en slechts als $\models [!p]\varphi$.

$$\models [!\varphi][!\varphi]S(\varphi) = 0.$$

Deze stelling zegt dat indien een gebeurtenis zich twee maal voordoet, het subject niet langer verrast zal zijn (ongeacht de intensiteit van verrassing na de eerste maal). We zagen dit effect reeds in het hierboven beschreven scenario: nadat Jan een tweede keer uit het raam heeft gekeken, is hij niet langer verrast dat het aan het sneeuwen is. Deze stelling zegt dus dat het *vluchtige* karakter van verrassing – waar binnen de psychologische literatuur reeds verschillende keren op gewezen is – ook binnen het logische systeem gevat kan worden.

Deze laatste stelling illustreert overigens ook een bijkomend voordeel van het dynamisch-logische systeem. Het is namelijk zo dat de eerdere, statische analyses van verrassing *niet* in staat zijn om het vluchtige karakter van verrassing te modelleren, en dus een belangrijke empirisch vastgestelde eigenschap van verrassing onverklaard moeten laten. In de dynamische analyse, daarentegen, kan dit vluchtige karakter wél verklaard worden, en het achterliggende formele bewijs (dat hier niet in detail besproken zal worden) leunt bovendien zeer dicht aan bij de informele verklaringen uit de psychologische literatuur. In vergelijking met de vroegere statische analyses zorgt het nieuwe dynamische systeem dus niet alleen voor de broodnodige *conceptuele* opheldering, maar is het bovendien ook adequater op *empirisch* vlak.

6. BESLUIT

In dit artikel heb ik de historische ontwikkeling van de epistemische logica beschreven, en hierbij bijzondere aandacht besteed aan de zogenaamde ‘dynamische wending’. Hoewel deze wending historisch gezien voornamelijk geïnspireerd was door computerwetenschappelijke en speltheoretische consideraties, heb ik geargumenteed dat zij wel degelijk ook relevant kan zijn op filosofisch vlak. De toepassing van dynamische logica’s op schijnbaar statische noties, stellingen, etc. past immers perfect binnen een wittgensteiniaanse opvatting over de rol van filosofie als conceptuele opheldering. Bovendien blijkt deze conceptuele opheldering vaak ook gepaard te gaan met een aantal andere, bijkomende voordelen, zoals een hogere mate van empirische adequaatheid. Ik heb deze argumentatie geconcretiseerd door de logica van publieke aankondigingen te presenteren, en te tonen hoe het psychologische fenomeen

‘verrassing’ op conceptueel en empirisch vruchtbare wijze binnen deze logica gemodelleerd kan worden.

Hoewel de notie van verrassing een aantal belangrijke filosofische implicaties heeft (zie voetnoten 34 en 35), kan men bezwaarlijk stellen dat zij tot de kernbegrippen van de filosofie behoort. Om het scepticisme van een aantal filosofen en filosofische logici ten aanzien van de dynamische wending in de epistemische logica helemaal weg te nemen, zal het wellicht nodig zijn om in toekomstig onderzoek aan te tonen dat dynamische analyses ook tot conceptuele opheldering kunnen leiden bij noties die wél tot de ‘harde kern’ van de filosofie behoren. Plausibele kandidaten zijn bijvoorbeeld de noties van type- en tokencausaliteit uit de analytische metafysica, maar de formele uitwerking hiervan staat momenteel nog in de kinderschoenen.⁴¹

SLEUTELWOORDEN: epistemische logica, dynamische wending, conceptuele opheldering, verrassing.

KEYWORDS: epistemic logic, dynamic turn, conceptual elucidation, surprise.

SUMMARY: *The Dynamic Turn in Epistemic Logic*

This article describes the historical development of epistemic logic, focusing on the dynamic turn that has taken place in the last few decades. Although this dynamic turn was mainly motivated by technical considerations in computer science and game theory, it is argued that it can also be relevant from a more philosophical perspective: the application of dynamic epistemic logics to analyze *prima facie* static notions, theorems, etc. fits perfectly in a Wittgensteinian approach to philosophy as conceptual elucidation. Furthermore, it turns out that these conceptual elucidations often lead to a number of other advantages, such as a higher degree of empirical adequacy. In order to illustrate this line of argumentation, the system of public announcement logic is presented, and it is shown how this system can be used to analyze the psychological phenomenon of surprise in a conceptually and empirically fruitful way.

⁴¹ Hartelijk dank aan Jan Heylen, Hans Smessaert, Margaux Smets en twee anonieme referenten voor hun commentaren op eerdere versies van dit artikel. De auteur is Postdoctoraal Onderzoeker van het Fonds Wetenschappelijk Onderzoek-Vlaanderen (FWO).